ON THE STABILITY AND ERGODICITY OF AN ADAPTIVE SCALING METROPOLIS ALGORITHM

MATTI VIHOLA

ABSTRACT. The stability and ergodicity properties of an adaptive random walk Metropolis algorithm are considered. The algorithm adjusts the scale of the symmetric proposal distribution continuously based on the observed acceptance probability. Unlike the previously proposed forms of this algorithm, the adapted scaling parameter is not constrained within a predefined compact interval. This makes the algorithm more generally applicable and 'automatic,' with two parameters less to be adjusted. A strong law of large numbers is shown to hold for functionals bounded on compact sets and growing at most exponentially as $\|x\| \to \infty$, assuming that the target density is smooth enough and has either compact support or super-exponentially decaying tails.

1. INTRODUCTION

Markov chain Monte Carlo (MCMC) is a general method often used to approximate integrals of the type

$$I := \int_{\mathbb{R}^d} f(x)\pi(x)\mathrm{d}x < \infty$$

where π is a probability density function [see, e.g., 8, 14, 17]. The method is based on a Markov chain $(X_n)_{n\geq 1}$ that can be simulated in practice, and for which $I_n := \sum_{k=1}^{n} f(X_k) \to I$ as $n \to \infty$. Such a chain can be constructed, for example, as follows. Assume q is a zero-mean Gaussian probability density, and let $X_1 \equiv x_1$ for some fixed point $x_1 \in \mathbb{R}^d$. For $n \geq 2$, recursively,

(S1) set $Y_n := X_{n-1} + \theta W_n$, where W_n are independent random vectors distributed according to q, and

(S2) with probability $\alpha_n := \min\{1, \pi(Y_n)/\pi(X_{n-1})\}$ the proposal is accepted and $X_n = Y_n$; otherwise the proposal is rejected and $X_n = X_{n-1}$.

For any positive scalar parameter θ, this symmetric random walk Metropolis algorithm is valid: $I_n \to I$ almost surely as $n \to \infty$ [e.g. 13, Theorem 1]. However, the

efficiency of the method, that is, the speed at which I_n converges to I, is crucially affected by the choice of θ. For too large θ, few proposals become accepted, and the chain mixes poorly. For too small θ, most of the proposals Y_n become accepted, but the steps $X_n - X_{n-1}$ are small, preventing good mixing. In fact, previous results indicate that the acceptance probability is closely related with the efficiency of the algorithm. The 'rule of thumb' is that the acceptance probability α_n should be on the average about 0.234 although this choice is not always optimal [7, 18, 19, 21]. In practice, such a θ is usually found by several trial runs, which can be laborious and time-consuming.

So called adaptive MCMC algorithms have gained popularity since the seminal work of Haario, Saksman, and Tamminen [10]. Several other such algorithms have been proposed after Andrieu and Robert [2] noticed the connection between Robbins-Monro stochastic approximation and adaptive MCMC [1, 3, 6, 15, 16]. The Adaptive Scaling Metropolis (ASM) algorithm considered in this paper optimises the scaling of the proposal distribution adaptively, based on the observed acceptance probability. Namely, in the step (S1) of the above algorithm, the constant θ is replaced with a random variable θ_{n-1} set initially to $\theta_1 > 0$ and for $n \geq 2$ defined recursively through

(S3) $\log \theta_n = \log \theta_{n-1} + \eta_n(\alpha_n - \alpha^*)$

where α^* is the desired mean acceptance probability, for example $\alpha^* = 0.234$, and $(\eta_n)_{n \geq 2}$ is a sequence of adaptation step sizes decaying to zero.

A similar random walk Metropolis algorithm with adaptive scaling was actually proposed over a decade ago by Gilks, Roberts and Sahu [9]. Their approach differed from the ASM approach so that the adaptation was performed only at particular regeneration times, which may occur infrequently or may be difficult to identify in practice. The ASM algorithm presented above has been proposed earlier by several authors [2, 6, 16], perhaps with a slightly different update formula (S3). The exact form of (S3) was used by [3, 5]. The crucial difference of the present paper compared to the earlier works is that the algorithm does not involve any additional constraints on θ_n.

This difference is chiefly theoretical advance, as discussed below. Therefore, no empirical studies of the performance of the ASM algorithm are included in the paper. It is nonetheless worth pointing out that since the ASM algorithm only adapts the scale of the proposal distribution, it is likely to be inefficient in certain situations. For example, if π is high-dimensional and possesses a strong correlation structure, the ASM approach is likely to be suboptimal. In such a situation, one could use the Adaptive Metropolis (AM) algorithm [10]. It has also been suggested to combine the AM algorithm with ASM [3, 5]; indeed, the analysis of the present paper can be used also in this setting (see Remarks 19 and 24).

It is not obvious that the ASM algorithm is valid, that is, $I_n \to I$. In fact, there are examples of continuously adapting MCMC schemes that destroy the correct ergodic properties [15]. Current ergodicity results on adaptive MCMC algorithms assume some 'uniform' behaviour for all the possible MCMC kernels [5, 6, 15]. In the context of the ASM algorithm, this essentially means that θ_n must be constrained

to a predefined set $[a, b]$ with some $0 < a \le b < \infty$. Alternatively, one can use a general reprojection technique with a sequence of such sets $[a_n, b_n]$ with $a_n \searrow 0$ and $b_n \nearrow \infty$ as proposed by Andrieu and Moulines [1], or stabilisation methods that modify the adaptation rule to ensure stable behaviour [3]. Such constraints and stabilisation structures are theoretically convenient, but may pose a problem for a practitioner. Good values for the constraint parameters may be difficult to choose without prior knowledge of the target distribution π. In the worst case, the values are chosen inappropriately and the algorithm is rendered useless in practice.

However, it is a common belief that many of the proposed adaptive MCMC algorithms are inherently stable and thereby do not require additional constraints or stabilisation structures. Indeed, there is considerable empirical evidence of the stability of several unconstrained algorithms, including the ASM approach. There are yet only few theoretical results, especially Saksman and Vihola [20] verifying the correct ergodic properties and the stability of the Adaptive Metropolis algorithm [10], provided the target distribution π has super-exponentially decaying tails with regular contours. These assumptions on π are close to those that ensure the geometric ergodicity of a non-adaptive random walk Metropolis algorithm [12]. The result in [20] does not assume an upper bound, but requires an explicit lower bound for the adapted covariance parameter. In the context of the ASM algorithm, this is analogous to constraining θ_n to the interval $[a, \infty)$, where $a > 0$.

The main results of this paper, formulated in the next section, show that the stability and ergodicity of the ASM algorithm can be verified under similar assumptions on the target distribution as in [20], without any modifications or constraints on the adaptation parameter $\theta_n \in (0, \infty)$. These are the first results that validate the correctness of a completely unconstrained, fully adaptive MCMC algorithm.

2. The Main Results

Throughout this section, suppose that the process $(X_n, \theta_n)_{n \ge 1}$ follows the ASM recursion (S1)–(S3) described in Section 1, the proposal density q is standard Gaussian and the step sizes are defined as $\eta_n := cn^{-\gamma}$ with some constants $c > 0$ and $\gamma \in (1/2, 1]$. Before stating the first ergodicity result, consider the following condition on the regularity of a collection of sets. Before that, recall that a C^1 domain in $\mathbb{R}^d$ is a domain whose boundary is locally a graph of a C^1 function.

Definition 1. Suppose that $\{A_i\}_{i \in I}$ is a collection of sets $A_i \subset \mathbb{R}^d$ each consisting of finitely many disjoint components that are closures of C^1 domains. Let $n_i(x)$ stand for the outer-pointing normal at x in the boundary ∂A_i. Then, $\{A_i\}_{i \in I}$ have *uniformly continuous normals* if for all $\epsilon > 0$ there is a $\delta > 0$ such that for any $i \in I$ it holds that $\|n_i(x) - n_i(y)\| \le \epsilon$ for all $x, y \in \partial A_i$ such that $\|x - y\| \le \delta$.

This definition essentially states that the boundaries ∂A_i must be regular enough to ensure that if one looks at ∂A_i at a small enough scale, it will look almost like a plane.

Theorem 2. *Assume π has a compact support $\mathbb{X} \subset \mathbb{R}^d$ and π is continuous, bounded and bounded away from zero on $\mathbb{X}$. Moreover, assume that the set $\mathbb{X}$ has a uniformly continuous normal (Definition 1). Then, for any $0 < \alpha^* < 1/2$ and a bounded function f, the strong law of large numbers holds, that is,*

$$(1) \qquad \frac{1}{n}\sum_{k=1}^{n} f(X_k) \xrightarrow{n\to\infty} \int_{\mathbb{R}^d} f(x)\pi(x)\mathrm{d}x \qquad almost\ surely.$$

Proof. Theorem 2 is a special case of Theorem 21 in Section 5. □

Let us consider next target distributions π with unbounded supports, satisfying the following conditions formulated in [20].

Assumption 3. The density π is bounded, bounded away from zero on compact sets, differentiable, and

$$(2) \qquad \lim_{r\to\infty} \sup_{\|x\|\geq r} \frac{x}{\|x\|^\rho} \cdot \nabla \log \pi(x) = -\infty$$

for some $\rho > 1$, where $\|\cdot\|$ stands for the Euclidean norm. Moreover, the contour normals satisfy

$$(3) \qquad \lim_{r\to\infty} \sup_{\|x\|\geq r} \frac{x}{\|x\|} \cdot \frac{\nabla \pi(x)}{\|\nabla \pi(x)\|} < 0.$$

This assumption is very near to the conditions introduced by Jarner and Hansen [12] to ensure the geometric ergodicity of a (non-adaptive) Metropolis algorithm, and considered by Andrieu and Moulines [1] in the context of adaptive MCMC. In particular, [1, 12] assume that π fulfils the contour regularity condition (3). Instead of (2), they assume a super-exponential decay on π,

$$\lim_{r\to\infty} \sup_{\|x\|\geq r} \frac{x}{\|x\|} \cdot \nabla \log \pi(x) = -\infty$$

which is only slightly more general than (2). See [12] for examples and discussion on the conditions.

Theorem 4. *Suppose π fulfils Assumption 3 and there is a $t_0 > 0$ such that the contour sets $\{L_t\}_{0<t\leq t_0}$ where $L_t := \{x \in \mathbb{R}^d : \pi(x) \geq t\}$ have uniformly continuous normals (Definition 1). Assume the function f is bounded on compact sets and grows at most exponentially, that is, there exist constants $M, \xi < \infty$ such that $|f(x)| \leq M \max\{1, e^{\xi\|x\|}\}$ for all $x \in \mathbb{R}^d$. Then, for any $0 < \alpha^* < 1/2$, the strong law of large numbers (1) holds.*

Proof. Theorem 4 is a special case of Theorem 23 in Section 5. □

Remark 5. For many practical target densities satisfying Assumption 3 the tail contours are (essentially) scaled copies of each other, in which case they have automatically uniformly continuous normals. This indicates that Theorem 4 is practically a counterpart of [20, Theorem 13] verifying the ergodicity of the Adaptive Metropolis algorithm.

Remark 6. The 'safe' values for the desired acceptance rate stipulated by Theorems 2 and 4 are $\alpha^* \in (0, 1/2)$. The values $[1/2, 1)$ are excluded due to technical reasons, in particular due to Proposition 12 establishing the lower bound for θ_n.[1] It is expected that Theorems 2 and 4 would hold assuming only $\alpha^* \in (0, 1)$, but this cannot be verified with the present technical approach. The range $\alpha^* \in (0, 1/2)$ is, however, often sufficient in practice, as the most commonly used values for a random walk Metropolis algorithms are probably $\alpha^* = 0.234$ and $\alpha^* = 0.44$, and it has been suggested that values $\alpha^* \in [0.1, 0.4]$ should work well in most cases [7, 16, 18, 19].

Remark 7. The results below hold for the above algorithmic setting, but allow some modifications. One can use a non-Gaussian proposal distribution q. In particular, the results hold for a heavy-tailed multivariate Student proposals. The step size sequence $(\eta_n)_{n\geq 2}$ can be selected quite freely; essentially, $(\eta_n)_{n\geq 2}$ must only be square-summable. Observe, however, that a sequence with $\sum_n \eta_n < \infty$ prevents efficient adaptation, as then θ_n is trivially bounded within $[a, b]$ for some $0 < a \leq b < \infty$.

The rest of the paper is organised as follows. Section 3 describes a general adaptive MCMC framework and a generalised version of the above described ASM algorithm within it. Section 4 develops stability results for this process. In particular, Corollary 14 ensures the stability of the sequence θ_n with the assumptions of Theorem 2, and Proposition 17 controls the growth of θ_n when π fulfils the conditions of Theorem 4. Once the stability results are obtained, the ergodicity is verified in Section 5 using the results in [20].

3. NOTATION AND FRAMEWORK

Consider first a general adaptive MCMC process evolving in a measurable space $\mathbb{X} \times \mathbb{S}$, where $\mathbb{X}$ is the space of the 'MCMC' chain $(X_n)_{n\geq 1}$ and $\mathbb{S}$ the space of the adaptation parameter $(S_n)_{n\geq 1}$. The process starts at some given $X_1 \equiv x_1 \in \mathbb{X}$ and $S_1 \equiv s_1 \in \mathbb{S}$, and for $n \geq 1$, follows the recursion[2]

$$(4) \qquad X_{n+1} = \begin{cases} Y_{n+1}, & \text{if } U_{n+1} \leq \alpha_{S_n}(X_n, Y_{n+1}) \\ X_n, & \text{otherwise} \end{cases}$$

$$(5) \qquad S_{n+1} = S_n + \eta_{n+1} H(S_n, X_n, Y_{n+1})$$

where the acceptance probability $\alpha_s : \mathbb{X} \times \mathbb{X} \to [0, 1]$ for each $s \in \mathbb{S}$, and $H : \mathbb{S} \times \mathbb{X} \times \mathbb{X} \to K_H$ is an adaptation function, with $K_H \subset \mathbb{R}$ compact. The σ-algebras $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$ are assumed to be such that the random variables U_n and Y_n are $\mathcal{F}_n$-measurable, U_{n+1} is independent on $\mathcal{F}_n$ and uniformly distributed on $[0, 1]$ and Y_{n+1} depends on $\mathcal{F}_n$ only via X_n and S_n. Namely, Y_n are distributed by the proposal density q_s so that $\mathbb{P}(Y_{n+1} \in A \mid \mathcal{F}_n) = \int_A q_{S_n}(X_n, y)\mathrm{d}y$. The sequence of non-negative step sizes η_n decays to zero.

---

[1]The S_n of Proposition 12 and θ_n are related by $\theta_n = e^{S_n}$

[2]The recursion of (5) can be considered as Robbins-Monro stochastic approximation; see [1, 2, 4] and references therein.

Let us consider next a generalisation of the ASM algorithm of Section 1. Let $\mathbb{S} = \mathbb{R}$ and define $\mathbb{X} \subset \mathbb{R}^d$ as the support of π. The family of proposal densities is defined as $q_s(x, y) := q_s(x - y)$ with

$$q_s(z) := [\phi(s)]^{-d} q([\phi(s)]^{-1} z)$$

where the template probability density q on $\mathbb{R}^d$ is symmetric, and the scaling function $\phi : \mathbb{R} \to (0, \infty)$ is increasing and surjective. To shed light on this definition, let Y be distributed according to q. Then, $\phi(s)Y$ is distributed according to q_s. In the context of the particular version of the algorithm described in Section 1, one has $\phi(s) = e^s$ and $S_n = \log \theta_n$. The acceptance probability is the Metropolis-Hastings ratio[3]

$$\alpha_s(x, y) := \alpha(x, y) := \min \left\{ 1, \frac{\pi(y)}{\pi(x)} \right\}.$$

The adaptation function H is defined as $H(s, x, y) := H(x, y) := \alpha(x, y) - \alpha^*$ where α^* is the constant desired acceptance rate, and the step sizes satisfy $\sum_{k=2}^{\infty} \eta_k = \infty$ and $\sum_{k=2}^{\infty} \eta_k^2 < \infty$.

Define the expected acceptance rate at $x \in \mathbb{X}$ with parameter $s \in \mathbb{S}$ as

$$\mathrm{acc}(x, s) := \int_{\mathbb{X}} \alpha(x, y) q_s(x - y) \mathrm{d}y.$$

Clearly, the adaptation rule decreases S_n whenever $\mathrm{acc}(X_n, S_n) < \alpha^*$, and vice versa. So, it is plausible that the algorithm would result in $S_n \to s^*$ such that $\mathrm{acc}(s^*) = \alpha^*$, where

$$\mathrm{acc}(s) := \int_{\mathbb{X}} \mathrm{acc}(x, s) \pi(x) \mathrm{d}x$$

is the expected acceptance rate over the target density π. In this paper, however, the convergence of S_n is not the main concern, but the stability of it, as it turns out to be crucial for the validity of the ASM algorithm.

The Metropolis transition kernel with a proposal density q_s is given as

$$(6) \qquad P_s(x, A) := \mathbb{1}_A(x) \int_{\mathbb{R}^d} [1 - \alpha(x, y)] q_s(x - y) \mathrm{d}y + \int_A \alpha(x, y) q_s(x - y) \mathrm{d}y$$

where $\mathbb{1}_A$ stands for the characteristic function of the set A. Using the kernels P_s, one can write (4) as $\mathbb{P}(X_{n+1} \in A \mid \mathcal{F}_n) = P_{S_n}(X_n, A)$. As usual, integration of a function f with respect to a transition kernel is denoted as

$$P_s f(x) := \int_{\mathbb{X}} f(y) P_s(x, \mathrm{d}y).$$

Let $V \geq 1$ be a function. The V-norm of a function f is defined as

$$\|f\|_V := \sup_x \frac{|f(x)|}{V(x)}.$$

The closed ball in $\mathbb{R}^d$ is written as $\overline{B}(x, r) := \{y \in \mathbb{R}^d : \|x - y\| \leq r\}$, and the distance of a point $x \in \mathbb{R}^d$ from the set $A \subset \mathbb{R}^d$ is denoted as $d(x, A) := \inf\{\|x - y\| : y \in A\}$.

---

[3]Note that Y_{n+1} may lie outside $\mathbb{X}$, but $(X_n)_{n \geq 1} \subset \mathbb{X}$ almost surely.

4. STABILITY

This section develops stability results, starting with a simple theorem on the general process given in Section 3. This theorem is auxiliary for the present paper, but may have applications with other adaptive MCMC algorithms of similar type.

Theorem 8. *Suppose $(X_n, S_n)_{n\geq 1}$ follow the general recursions (4) and (5), and the step sizes satisfy $\sum_{n=1}^{\infty} \eta_n^2 < \infty$.*

(i) If there is a constant $a < \infty$ such that for all $n \geq 1$

$$\mathbb{E}\left[H(S_n, X_n, Y_{n+1}) \mid \mathcal{F}_n\right] \leq 0 \qquad \text{whenever } S_n \geq a,$$

then $\limsup_{n\to\infty} S_n < \infty$ a.s.

(ii) If also $\sum_n \eta_n = \infty$ and there is a non-decreasing sequence of constants $(a_n)_{n\geq 1} \subset \mathbb{R}$ such that

$$\mathbb{E}\left[H(S_n, X_n, Y_{n+1}) \mid \mathcal{F}_n\right] \leq b \qquad \text{whenever } S_n \geq a_n$$

for some $b < 0$, then $\limsup_{n\to\infty}(S_n - a_n) \leq 0$ a.s.

Proof. Let $W_n := H(S_{n-1}, X_{n-1}, Y_n)\mathbb{1}_{\{S_{n-1}\geq a\}}$ for $n \geq 2$, and define the martingale $(M_n, \mathcal{F}_n)_{n\geq 1}$ by setting $M_1 := 0$, and $M_n := \sum_{k=2}^{n} dM_k$ for $n \geq 2$ with the differences $dM_n := \eta_n(W_n - \mathbb{E}\left[W_n \mid \mathcal{F}_{n-1}\right])$. Clearly,

$$\sum_{k=2}^{\infty} \mathbb{E}\left[dM_k^2 \mid \mathcal{F}_{k-1}\right] \leq 4c^2 \sum_{k=2}^{\infty} \eta_k^2 < \infty$$

where $c = \sup_{x\in K_H} |x|$. This implies that M_n converges to an a.s. finite limit M_∞ [e.g. 11, Theorem 2.15].

Let $(\tau_k)_{k\geq 1}$ be the exit times of S_n from $(-\infty, a)$, defined as $\tau_k := \inf\{n > \tau_{k-1} : S_n \geq a, S_{n-1} < a\}$ using the conventions $\tau_0 = 0$, $S_0 < a$, and $\inf \emptyset = \infty$. Define also the latest exit from $(-\infty, a)$ by $\sigma_n := \sup\{\tau_k : k \geq 1, \tau_k \leq n\}$. Whenever $S_n \geq a$, one can write $S_n = S_{\sigma_n} + (M_n - M_{\sigma_n}) + Z_{\sigma_n, n}$ where

$$Z_{m,n} := \sum_{k=m+1}^{n} \eta_n \mathbb{E}\left[W_n \mid \mathcal{F}_{n-1}\right] \leq 0$$

by assumption. In this case,

$$(7) \qquad S_n \leq S_{\sigma_n} + (M_n - M_{\sigma_n}) \leq \max\{S_1, a + c\eta_{\sigma_n}\} + 2\sup_{k\geq 1}|M_k| \leq C$$

where C is a.s. finite. If $S_n < a$ the claim is trivial and (i) holds.

Assume then (ii). If $S_n < a_n$ for all n greater than some $N_1(\omega) < \infty$, the claim is trivial. Suppose then that $S_n \geq a_n$ infinitely often. Define $(\tau_k)_{k\geq 1}$ as the exit times of S_n from $(-\infty, a_n)$ as above. The times τ_k must be a.s. finite in this case (and S_n returns to $(-\infty, a_n)$ infinitely often), for suppose the contrary: then the last exit times σ_n are bounded by some $\sigma_n \leq \sigma < \infty$, and for $n \geq \sigma$ one may write

$$S_n = S_\sigma + (M_n - M_\sigma) + Z_{\sigma,n} \leq C_\sigma + Z_{\sigma,n}$$

where M_n and $Z_{n,m}$ are defined as above, but using the random variables $W_n := H(S_{n-1}, X_{n-1}, Y_n)\mathbb{1}_{\{S_{n-1}\geq a_{n-1}\}}$, and the random variable C_σ is a.s. finite as in (7). Now, $Z_{\sigma,n} \to -\infty$ a.s. as $n \to \infty$, so $S_n < a_n$ a.s. for sufficiently large n, which is a contradiction.

Fix an $\epsilon > 0$ and let $N_0 = N_0(\omega, \epsilon)$ be such that for all $n \geq N_0$, it holds that $c\eta_{\sigma_n} \leq \epsilon/3$ and that $|M_k - M_\infty| \leq \epsilon/3$ a.s. for all $k \geq \sigma_n$. The claim follows from the estimate

$$S_n \leq S_{\sigma_n} + (M_n - M_{\sigma_n}) = S_{\sigma_n - 1} + \eta_{\sigma_n}H(S_{\sigma_n - 1}, X_{\sigma_n - 1}, Y_{\sigma_n}) + (M_n - M_{\sigma_n})$$
$$\leq a_{\sigma_n} + \epsilon/3 + |M_n - M_\infty| + |M_\infty - M_{\sigma_n}| \leq a_n + \epsilon$$

for all $n \geq N_0$. □

Remark 9. Theorem 8 generalises for an unbounded adaptation function H under suitable additional assumptions. For example, assuming

$$\limsup_{n\to\infty} |\eta_{n+1}H(S_n, X_n, Y_{n+1})| = 0 \quad \text{and}$$

$$\sum_{k=1}^{\infty} \eta_{n+1}^2 \mathbb{E}\left[H^2(S_n, X_n, Y_{n+1}) \mid \mathcal{F}_n \right] < \infty$$

hold almost surely, the proof applies with obvious changes. Moreover, the function H may depend additionally on U_{n+1} (or X_{n+1}).

Hereafter, consider the adaptive scaling Metropolis (ASM) algorithm described in Section 3. One can give simple conditions under which the result of Theorem 8 applies. This is due to the fact that one can write

$$\mathbb{E}\left[H(S_n, X_n, Y_{n+1}) \mid \mathcal{F}_n \right] = \mathrm{acc}(X_n, S_n) - \alpha^*,$$

so in light of Theorem 8, it is sufficient to find out when $\mathrm{acc}(x, s)$ is below or above α^*.

Proposition 10. *Assume π is supported on a compact set $\mathbb{X} \subset \mathbb{R}^d$ and $\alpha^* > 0$. Then, there is $b < 0$ and $a \in \mathbb{R}$ such that*

(8) $$\mathbb{E}\left[H(S_n, X_n, Y_{n+1}) \mid \mathcal{F}_n \right] \leq b \qquad \text{whenever } S_n \geq a.$$

Proof. Without loss of generality, one can assume $0 \in \mathbb{X}$. Let $\epsilon > 0$ be sufficiently small so that $\int_{\overline{B}(0,\epsilon)} q(z)\mathrm{d}z \leq \alpha^*/2$, and let a be sufficiently large so that $\phi(s) \geq \mathrm{diam}(\mathbb{X})\epsilon^{-1}$ for all $s \geq a$. Then, for all $x \in \mathbb{X}$,

$$\int_{\mathbb{X}} \alpha(x,y)q_s(x-y)\mathrm{d}y \leq \int_{\mathbb{X}} [\phi(s)]^{-d}q([\phi(s)]^{-1}z)\mathrm{d}z = \int_{[\phi(s)]^{-1}\mathbb{X}} q(u)\mathrm{d}u$$

$$\leq \int_{\overline{B}(0,\epsilon)} q(u)\mathrm{d}u \leq \frac{\alpha^*}{2}.$$

That is, (8) holds with $b = -\alpha^*/2 < 0$, whenever $s \geq a$. □

Before stating the next result bounding the conditional expectation to the opposite direction, let us consider a condition on the tails of π.

Assumption 11. There is a $\lambda > 0$ such that $L_\lambda := \{y \in \mathbb{R}^d : \pi(y) \geq \lambda\}$ is compact and π is continuous on L_λ. Moreover, the sets in the collection $\{L_t\}_{0 < t \leq \lambda}$ have uniformly continuous normals (Definition 1).

Proposition 12. *Suppose the target density π satisfies Assumption 11. Then, for any $\alpha^* < 1/2$, there are $a \in \mathbb{R}$ and $b > 0$ such that*

$$(9) \qquad \mathbb{E}\left[H(S_n, X_n, Y_{n+1}) \mid \mathcal{F}_n \right] \geq b \qquad \text{whenever } S_n \leq a.$$

Before giving the proof of Proposition 12, let us outline the simple intuition behind it. For all s small enough, the mass of q_s is essentially concentrated on a small ball $\overline{B}(0, \epsilon)$. If one looks the target π only on $\overline{B}(x, \epsilon)$, there are basically two alternatives. The first one is that π is approximately constant on that small ball and $\mathrm{acc}(x, s) \approx 1$. The second alternative is that π decreases very rapidly to *one* direction, in which case the set $\{y : \pi(y) \geq \pi(x)\}$ looks like a half-space on the ball $\overline{B}(x, \epsilon)$, and $\mathrm{acc}(x, s) \gtrsim 1/2$.

Let us start with a lemma on this 'half-space approximation.'

Lemma 13. *Suppose that the sets $\{A_i\}_{i \in I}$ with $A_i \subset \mathbb{R}^d$ have uniformly continuous normals (Definition 1). Then, for any $\epsilon > 0$, there is a $\delta > 0$ such that for any $i \in I$, any $x \in A_i$ and any $0 < r \leq \delta$, there is a half-space T such that $\overline{B}(x, r) \cap T \subset \overline{B}(x, r) \cap A_i$, and the distance $d(x, T) \leq \epsilon r$.*

The claim is geometrically evident. The technical verification is given in Appendix A.

Proof of Proposition 12. Fix an $\epsilon^* \in (0, 1)$ and let $M \geq 1$ be sufficiently large so that

$$(10) \qquad \int_{\overline{B}(0, \phi(s)M)} q_s(z)\mathrm{d}z = \int_{\overline{B}(0, M)} q(z)\mathrm{d}z \geq 1 - \epsilon^*$$

and for any plane P, it holds that

$$(11) \qquad \int_{\{d(z, P) \leq \phi(s)M^{-1}\}} q_s(z)\mathrm{d}z = \int_{\{d(z, P) \leq M^{-1}\}} q(z)\mathrm{d}z \leq \epsilon^*.$$

By compactness of L_λ and positivity of π one can find $\delta_1 > 0$ such that for all $x, y \in L_\lambda$ with $\|x - y\| \leq \delta_1$, it holds that $|\log \pi(x) - \log \pi(y)| \leq \epsilon^*$ so that

$$1 - \alpha(x, y) = e^0 - e^{\min\{0, \log \pi(y) - \log \pi(x)\}} \leq |\log \pi(y) - \log \pi(x)| \leq \epsilon^*.$$

Let $\delta_2 > 0$ be sufficiently small to satisfy Lemma 13 with the choice $\epsilon = M^{-2}$.

Choose a small enough $a \in \mathbb{R}$ so that $\phi(a)M \leq \min\{\delta_1, \delta_2\}$. Let $s \leq a$, denote $r_s := \phi(s)M$, and write for any $x \in L_\lambda$

$$\int_{\mathbb{X}} \alpha(x, y) q_s(x - y)\mathrm{d}y \geq \int_{\overline{B}(x, r_s) \cap L_\lambda} \alpha(x, y) q_s(x - y)\mathrm{d}y$$

$$\geq (1 - \epsilon^*) \int_{\overline{B}(x, r_s) \cap L_\lambda} q_s(x - y)\mathrm{d}y$$

since $r_s \leq \delta_1$. Denote by T the half-space from Lemma 13, such that $\overline{B}(x, r_s) \cap T \subset \overline{B}(x, r_s) \cap L_\lambda$ and the distance $d(x, T) \leq M^{-2} r_s$. One obtains

$$
\begin{aligned}
\int_{\mathbb{X}} \alpha(x, y) q_s(x - y) \mathrm{d}y &\geq (1 - \epsilon^*) \int_{\overline{B}(x, r_s) \cap T} q_s(x - y) \mathrm{d}y \\
&\geq (1 - \epsilon^*) \int_{\overline{B}(x, r_s) \cap \tilde{T}} q_s(x - y) \mathrm{d}y - \int_{\{d(y, P) \leq M^{-2} r_s\}} q_s(x - y) \mathrm{d}y \\
&\geq \frac{1}{2}(1 - \epsilon^*)^2 - \epsilon^*
\end{aligned}
$$

where $\tilde{T}$ is the half-space with the boundary plane P parallel to the boundary of T, and passing through x. The last inequality follows from (10) with the symmetry of q_s and (11), respectively. The same estimate holds for any $x \in L_t$ with $t > 0$.

To conclude,

$$
\mathrm{acc}(x, s) = \int_{\mathbb{X}} \alpha(x, y) q_s(x - y) \mathrm{d}y \geq \frac{1}{2} - \frac{1}{2}\left(\frac{1}{2} - \alpha^*\right)
$$

for all $x \in \mathbb{X}$ and for any $\alpha^* < 1/2$ by selecting $\epsilon^* = \epsilon^*(\alpha^*) > 0$ to be sufficiently small, implying (9) with $b = (1/2 - \alpha^*)/2 > 0$. $\square$

As an easy corollary of the propositions above, one establishes the stability of the ASM process.

Corollary 14. *Assume the target density π is compactly supported, and satisfies Assumption 11. Then, for the ASM process $(X_n, S_n)_{n \geq 1}$ with any $0 < \alpha^* < 1/2$, there are a.s. finite A_1 and A_2 such that*

$$
(12) \qquad\qquad\qquad\qquad A_1 \leq S_n \leq A_2
$$

for all $n \geq 1$.

Proof. The conditions of Propositions 10 and 12 are satisfied, so there are constants $-\infty < a_1 < a_2 < \infty$ and $b < 0$ such that

$$
\begin{aligned}
\mathbb{E}\left[H(S_n, X_n, Y_{n+1}) \mid \mathcal{F}_n\right] \leq b \qquad &\text{whenever} \qquad S_n \geq a_2, \\
\mathbb{E}\left[H(S_n, X_n, Y_{n+1}) \mid \mathcal{F}_n\right] \geq -b \qquad &\text{whenever} \qquad S_n \leq a_1.
\end{aligned}
$$

Theorem 8 applied to $-S_n$ and S_n guarantees that $a_1 \leq \liminf_{n \to \infty} S_n$ and $\limsup_{n \to \infty} S_n \leq a_2$, respectively, from which one obtains a.s. finite A_1 and A_2 for which (12) holds. $\square$

The rest of this section considers targets π with an unbounded support. Under a suitably regular π, it is shown that the growth of S_n can be controlled. To start with, consider the following properties for the scaling function ϕ and the template proposal distribution q.

Assumption 15. The scaling function ϕ is piecewise differentiable and there are constants $h, c > 0$ and $\kappa \geq 1$ such that

$$
\phi'(x + \xi) \leq c \max\{1, \phi^\kappa(x)\}
$$

for all $x \in \mathbb{R}$ and all $0 \leq \xi \leq h$.

Assumption 15 is not restrictive, and it clearly holds for any polynomial or exponential ϕ.

Assumption 16. The template proposal density q can be written as $q(z) = \hat{q}(\|\Sigma^{-1}z\|)$ where $\Sigma \in \mathbb{R}^{d \times d}$ is a symmetric and positive definite matrix and $\hat{q} : [0, \infty) \to (0, \infty)$ is a bounded, decreasing and differentiable function. Moreover, there is an $\epsilon^* > 0$ and $0 \leq a < b < \infty$ such that for all $0 \leq \epsilon \leq \epsilon^*$, the following bounds hold for the derivative of $\hat{q}$

$$\hat{q}'(x) - 2\hat{q}'(x + \epsilon) \geq c_1, \qquad \text{for all} \qquad a \leq x \leq b,$$

$$\int_0^\infty \min\{0, \hat{q}'(x) - 2\hat{q}'(x + \epsilon)\}\mathrm{d}x \geq -c_2 e^{-c_3 \epsilon^{-1}}$$

with some constants $c_1, c_2, c_3 > 0$.

Assumption 16 stipulates that q is elliptically symmetric and the contours of q have main axes proportional to the eigenvalues of Σ. Moreover, the decay rate of q along any ray is determined by $\hat{q}$ satisfying the technical bounds. Lemma 28 in Appendix B shows that Assumption 16 holds for Gaussian and Student distributions q.

The following estimate for the at most polynomial growth of $\phi(S_n)$ is crucial for the ergodicity result obtained in Theorem 23.

Proposition 17. *Suppose π satisfies Assumptions 3 and 11. Suppose also that the scaling function ϕ satisfies Assumption 15 and the template density q fulfils Assumption 16. Then, for the ASM process $(X_n, S_n)_{n \geq 1}$ with $0 < \alpha^* < 1/2$, and for any $\beta > 0$, there is an a.s. positive $\Theta_1 = \Theta_1(\omega)$ and an a.s. finite $\Theta_2 = \Theta_2(\omega, \beta)$ such that*

$$\Theta_1 \leq \phi(S_n) \leq \Theta_2 n^\beta.$$

Before the proof, let us consider an estimate of $\mathrm{acc}(x, s)$ depending on both x and s.

Lemma 18. *Assume π satisfies Assumption 3. Then, for any $\epsilon > 0$, there is a constant $c = c(\epsilon) \geq 1$ such that $\mathrm{acc}(x, s) \leq \epsilon$ for all $\phi(s) \geq c \max\{1, \|x\|\}$.*

Proof. Let $r_1 \geq 1$ be sufficiently large so that for some $\gamma > 0$ it holds that $\frac{x}{\|x\|} \cdot \frac{\nabla \pi(x)}{\|\nabla \pi(x)\|} < -\gamma$ and $\frac{x}{\|x\|^\rho} \cdot \nabla \log \pi(x) < -\gamma$ for all $\|x\| \geq r_1$. Increase r_1, if necessary, so that for any $\|x\| \geq r_1$ one can write $L_{\pi(x)} = \{y : \pi(y) \geq \pi(x)\} = \{ru : u \in S^d, 0 \leq r \leq g(u)\}$ where $S^d := \{u \in \mathbb{R}^d : \|u\| = 1\}$ is the unit sphere and the function $g : S^d \to (0, \infty)$ parameterises the boundary of $L_{\pi(x)}$. Notice also that the contour normal condition implies the existence of an $M \geq 1$ such that $L_{\pi(x)} \subset \overline{B}(0, M\|x\|)$ for all $\|x\| \geq r_1$.

Write for $\|x\| \geq r_2 := Mr_1$

$$
\mathrm{acc}(x, s) = \int_{\mathbb{R}^d} \alpha(x, y) q_s(x - y) \mathrm{d}y
$$

$$
\leq \int_{\{d(y, L_{\pi(x)}) \leq \|x\|\}} q_s(x - y) \mathrm{d}y + \sup_{y \in \mathbb{R}^d} q_s(x - y) \int_{\{d(y, L_{\pi(x)}) > \|x\|\}} \alpha(x, y) \mathrm{d}y.
$$

The first term can be estimated from above by

$$
\int_{\overline{B}(0, M\|x\| + \|x\|)} q_s(x - y) \mathrm{d}y \leq \int_{\overline{B}(0, (M+2)\|x\|)} q_s(z) \mathrm{d}z = \int_{\overline{B}(0, r(s, x))} q(u) \mathrm{d}u \leq \frac{\epsilon}{2}
$$

whenever $r(s, x) := [\phi(s)]^{-1}(M + 2)\|x\| \leq \epsilon^*$ for some small enough $\epsilon^* = \epsilon^*(\epsilon) > 0$, as in the proof of Proposition 10.

For the latter term, notice that

$$
\sup_{y \in \mathbb{R}^d} q_s(x - y) = [\phi(s)]^{-d} \sup_{z \in \mathbb{R}^d} q(z) \leq c_1 [\phi(s)]^{-d}.
$$

The integral can be estimated by polar integration as

$$
\int_{\{d(y, L_{\pi(x)}) > \|x\|\}} \alpha(x, y) \mathrm{d}y \leq c_d \sup_{u \in S^d} \int_{r > g(u) + \|x\|}^{\infty} r^{d-1} e^{\log \pi(ru) - \log \pi(g(u)u)} \mathrm{d}r
$$

where c_d is the surface measure of the sphere S^d. Since $\|x\| \geq r_2$, one has that $g(u) \geq r_1 \geq 1$, and from the gradient decay condition, one obtains that for $r > g(u) + 1$

$$
\log \pi(ru) - \log \pi(g(u)u) = \int_{g(u)}^{r} \frac{tu}{\|tu\|} \cdot \nabla \log \pi(tu) \mathrm{d}t \leq -\gamma \int_{g(u)}^{r} t^{\rho-1} \mathrm{d}t
$$

$$
\leq -\gamma g(u)^{\rho-1}[r - g(u)]
$$

from which

$$
\int_{r > g(u) + \|x\|}^{\infty} r^{d-1} e^{\log \pi(ru) - \log \pi(g(u)u)} \mathrm{d}r
$$

$$
\leq \int_0^{\infty} e^{-\frac{\gamma w}{2}} \mathrm{d}w \sup_{r > g(u) + \|x\|} r^{d-1} e^{-\frac{\gamma}{2} g(u)^{\rho-1}[r - g(u)]}.
$$

Consequently,

$$
\int_{\{d(y, L_{\pi(x)}) > \|x\|\}} \alpha(x, y) \mathrm{d}y
$$

$$
\leq c_d \frac{2}{\gamma} \sup_{\tilde{g} \geq 1, \tilde{r} > 1} \exp \left[(d - 1) \log(\tilde{g} + \tilde{r}) - \frac{\gamma}{2} \tilde{g}^{\rho-1} \tilde{r} \right] \leq c_2
$$

with a finite constant c_2 whenever $\|x\| \geq r_2$.

To sum up, there is a $c_3 > 0$ such that for any $\|x\| \geq r_2$ and any s satisfying

$$
\phi(s) \geq c_3 \max\{1, \|x\|\} \geq \max \left\{ \left(c_1 c_2 \frac{2}{\epsilon} \right)^{1/d}, \frac{(M + 2)\|x\|}{\epsilon^*} \right\},
$$

it holds that $\mathrm{acc}(x,s) \leq \epsilon$. For any $\|x\| < r_2$ there is a $r_2 \leq \|x_0\| \leq Mr_2$ such that $\pi(x_0) \leq \pi(x)$. Consequently, $\alpha(x,y) \leq \alpha(x_0,y)$ for all $y \in \mathbb{R}^d$ and therefore

$$\mathrm{acc}(x,s) \leq \int_{\{d(y,L_{\pi(x_0)}) \leq \|x_0\|\}} q_s(x-y)\mathrm{d}y$$
$$+ \sup_{y \in \mathbb{R}^d} q_s(x-y) \int_{\{d(y,L_{\pi(x)}) > \|x_0\|\}} \alpha(x_0,y)\mathrm{d}y.$$

Repeating the above arguments, there is a finite constant c_4 such that $\mathrm{acc}(x,s) \leq \epsilon$ whenever $\phi(s) \geq c_4 \max\{1, \|x\|\}$. $\qquad\square$

Having Lemma 18 and the lower bound from Proposition 12, the proof of Proposition 17 can be obtained by applying the growth condition on $\|X_n\|$ established in [20].

Proof of Proposition 17. Proposition 12 applied with Theorem 8 for $-S_n$ gives an a.s. finite A_1 such that $A_1 \leq S_n$. Since $\phi > 0$ is increasing, the variable $\Theta_1 := \phi(A_1)$ is a.s. positive, showing the lower bound.

To check the polynomial growth condition for $\phi(S_n)$, it is first verified that $\|X_n\|$ grows at most polynomially. Fix an $\epsilon > 0$ and let $\theta_1 = \theta_1(\epsilon) > 0$ and $a_1 = a_1(\epsilon) \in \mathbb{R}$ be such that $\theta_1 = \phi(a_1)$, and that $\mathbb{P}(B_1) \geq 1-\epsilon$, with $B_1 := \{\Theta_1 \geq \theta_1\} = \{A_1 \geq a_1\}$. Let $V(x) := c_\pi \pi^{-1/2}(x)$, where the constant $c_\pi := [\sup_x \pi(x)]^{1/2}$ ensures that $V \geq 1$. Proposition 26 in Appendix B shows that the drift inequality

(13) $$P_s V(x) \leq V(x) + b$$

holds for all $\phi(s) \geq \theta_1 > 0$ with some $b = b(\theta_1) < \infty$. Construct an auxiliary process $(X_n', S_n')_{n\geq 1}$ coinciding with $(X_n, S_n)_{n\geq 1}$ in B_1 by setting $(X_n', S_n') = (X_{\tau_n}, S_{\tau_n})$ where the stopping times τ_n are defined as

$$\tau_n := \begin{cases} n, & \text{if } \phi(S_k) \geq \theta_1 \text{ for all } 1 \leq k \leq n \\ \inf\{1 \leq k \leq n-1 : \phi(S_{k+1}) < \theta_1\}, & \text{otherwise.} \end{cases}$$

Having the inequality (13), set $\beta' = \kappa^{-1}\beta$ where the constant $\kappa \geq 1$ is from Assumption 15 and use Proposition 10 of [20] to obtain the bound $\|X_n'\| \leq \Theta_\epsilon n^{\beta'}$ for some a.s. finite Θ_ϵ. The $\epsilon > 0$ was arbitrary, so one can let $\epsilon \to 0$ and obtain an a.s. finite Θ such that $\|X_n\| \leq \Theta n^{\beta'}$. Applying Lemma 18, one obtains that $\mathrm{acc}(X_n, S_n) \leq \alpha^*/2$ whenever $\phi(S_n) \geq \Theta' n^{\beta'}$ with $\Theta' := c_1 \max\{1, \Theta\}$.

Fix again an $\epsilon > 0$ and let $\theta_2 = \theta_2(\epsilon) < \infty$ be such that $\mathbb{P}(B_2) \geq 1-\epsilon$ where $B_2 := \{\Theta' \leq \theta_2\}$. Construct an auxiliary process $(X_n', S_n')_{n\geq 1}$ coinciding with $(X_n, S_n)_{n\geq 1}$ in B_2 by stopping the process if $\phi(S_k) > \theta_2 k^{\beta'}$ as in the construction above. Theorem 8 ensures that

$$\limsup_{n\to\infty}[S_n' - \tilde{a}_n] \leq 0$$

where $\tilde{a}_n$ are defined so that $\phi(\tilde{a}_n) = \theta_2 n^{\beta'}$. That is, $S_n' \leq \tilde{a}_n + E_n$ with $E_n \to 0$ almost surely. Consider Assumption 15 and take N_0 so large that $E_n < h$ for all

$n \geq N_0$. Then, $\phi(x+h) = \phi(x) + h\phi'(x+\xi)$ for some $0 \leq \xi \leq h$, and hence $\phi(x+h) \leq c_2 \max\{1, \phi(x)^\kappa\}$. For $n \geq N_0$, one has

$$\phi(S'_n) \leq \phi(\tilde{a}_n + E_n) \leq c_2 \max\{1, \phi(\tilde{a}_n)^\kappa\} = c_2 \max\{1, \theta_2^\kappa n^{\kappa\beta'}\} \leq \theta_2' n^\beta$$

for some finite θ_2'. Summing up, there is an a.s. finite Θ_2' such that

$$\phi(S'_n) \leq \Theta_2' n^\beta$$

on B_2. Finally, letting $\epsilon \to 0$, one can find an a.s. finite Θ_2 such that $\phi(S_n) \leq \Theta_2 n^\beta$. $\qquad\square$

Remark 19. It is possible to obtain Corollary 14 and Proposition 17 when using the ASM algorithm within some other adaptation framework. For example, ASM can be combined with the Adaptive Metropolis algorithm as suggested in [5] and [3]. In particular, one could assume that there is another ($\mathcal{F}_n$-measurable) parameter $\tilde{S}_n$ in addition to S_n, so that $Y_{n+1} \sim q_{S_n, \tilde{S}_n}(X_n, \cdot)$ with

$$q_{s,\tilde{s}}(x,y) := q_{s,\tilde{s}}(x-y) := [\phi(s)]^{-d} q_{\tilde{s}}([\phi(s)]^{-1}(x-y))$$

where $\{q_{\tilde{s}}\}_{\tilde{s} \in \tilde{\mathbb{S}}}$ is a suitably 'uniform' family of symmetric probability densities. If there is an integrable function q^+ such that $q_{\tilde{s}} \leq q^+$ for all $\tilde{s} \in \tilde{\mathbb{S}}$ then Propositions 10 and 12 can be verified to hold, implying Corollary 14. Moreover, if $\tilde{\mathbb{S}}$ is a subset of positive definite $d \times d$ matrices with eigenvalues bounded away from zero and infinity and $q_{\tilde{s}}(z) = \det(\tilde{s})^{-1} \hat{q}(\|\tilde{s}^{-1} z\|)$ with $\hat{q}$ satisfying Assumption 16, then Proposition 17 can be shown to hold.

5. ERGODICITY

In Section 4, the stability or controlled growth of the ASM process was established under certain conditions. This section employs these results to prove strong laws of large numbers for the ASM process, relying on the results introduced in [20]. For this purpose, consider the following alternative theoretical adaptation introduced in [20], that applies to a sequence of restriction sets $K_1 \subset K_2 \subset \cdots \subset K_n \subset \mathbb{S}$.

Assume $(\tilde{X}_n, \tilde{S}_n)_{n \geq 1}$ follow the adaptation framework described in Section 3, with $(\tilde{Y}_n)_{n \geq 2}$ defined also as in (4). Assume $\tilde{S}_1 \equiv \tilde{s}_1 \in K_1$ and instead of (5) let $(\tilde{S}_n)_{n \geq 1}$ follow the 'truncated' recursion

$$(14) \qquad \tilde{S}_{n+1} = \sigma_{n+1}\big(\tilde{S}_n, \eta_{n+1} H(\tilde{S}_n, \tilde{X}_n, \tilde{Y}_{n+1})\big)$$

where the restriction functions $\sigma_n : \mathbb{S} \times \mathbb{S} \to \mathbb{S}$ are defined as

$$\sigma_n(s, s') := \begin{cases} s + s', & \text{if } s + s' \in K_n \\ s, & \text{otherwise.} \end{cases}$$

That is, σ_n ensures that $\tilde{S}_n \in K_n$ for all $n \geq 1$. Observe that such a 'truncated process' can be constructed using an 'original process' $(X_n, S_n)_{n \geq 1}$ and $(Y_n, U_n)_{n \geq 2}$ following (4) and (5), and so that the two processes coincide in the set $\cap_{n=1}^{\infty}\{S_n \in K_n\}$.

Before stating an ergodicity result for this truncated chain, four technical assumptions are listed, which must hold for some constants $c \geq 1$ and $\beta \geq 0$.

(A1) For all measurable $A \subset \mathbb{X}$, it holds that $\mathbb{P}(\tilde{X}_{n+1} \in A \mid \mathcal{F}_n) = P_{\tilde{S}_n}(\tilde{X}_n, A)$ almost surely, and for each $s \in \mathbb{S}$, the transition probability P_s has π as the unique invariant distribution.

(A2) For each $n \geq 1$, the following uniform drift and minorisation conditions hold for all $s \in K_n$, for all $x \in \mathbb{X}$ and all measurable $A \subset \mathbb{X}$

$$P_s V(x) \leq \lambda_n V(x) + b_n \mathbb{1}_{C_n}(x)$$
$$P_s(x, A) \geq \delta_n \mathbb{1}_{C_n}(x) \nu_s(A)$$

where $C_n \subset \mathbb{X}$ is a subset (a minorisation set), $V : \mathbb{X} \to [1, \infty)$ is a drift function such that $\sup_{x \in C_n} V(x) \leq b_n$ and ν_s is a probability measure on $\mathbb{X}$ concentrated on C_n. Furthermore, the constants $\lambda_n \in (0, 1)$ and $b_n \in (0, \infty)$ are increasing, $\delta_n \in (0, 1]$ is decreasing with respect to n and they are polynomially bounded so that

$$\max\{(1 - \lambda_n)^{-1}, \delta_n^{-1}, b_n\} \leq cn^\beta.$$

(A3) For all $n \geq 1$ and any $r \in (0, 1]$, there is $c' = c'(r) \geq 1$ such that for all $s, s' \in K_n$,

$$\|P_s f - P_{s'} f\|_{V^r} \leq c' n^\beta \|f\|_{V^r} |s - s'|.$$

(A4) The inequality $|H(\tilde{S}_n, \tilde{X}_n, \tilde{Y}_{n+1})| \leq cn^\beta$ holds almost surely.

Theorem 20. *Assume (A1)–(A4) hold and let f be a function with $\|f\|_{V^\gamma} < \infty$ for some $\gamma \in (0, 1)$. Assume $\beta < \kappa_*^{-1} \min\{1/2, 1 - \gamma\}$ and $\sum_{k=1}^\infty k^{\kappa_* \beta - 1} \eta_k < \infty$ where $\kappa_* \geq 1$ is an independent constant. Then,*

$$(15) \qquad \frac{1}{n} \sum_{k=1}^n f(\tilde{X}_k) \xrightarrow{n \to \infty} \int_{\mathbb{X}} f(x) \pi(x) \mathrm{d}x \quad \text{almost surely.}$$

Proof. This theorem is a straightforward modification of Theorem 2 in [20]. In particular, the assumption (A4) here is slightly simpler than assumption (A4) in [20] and the changes required for the proof are obvious. $\square$

The following first main result considers the case of compactly supported π.

Theorem 21. *Suppose π has a compact support $\mathbb{X} \subset \mathbb{R}^d$ and π is continuous, bounded and bounded away from zero on $\mathbb{X}$. Moreover, assume that the set $\mathbb{X}$ has a uniformly continuous normal (Definition 1) and the template proposal density q satisfies Assumption 16. Then, for the ASM process $(X_n, S_n)_{n \geq 1}$ with any $0 < \alpha^* < 1/2$ and a bounded function f, the strong law of large numbers holds, that is,*

$$(16) \qquad \frac{1}{n} \sum_{k=1}^n f(X_k) \xrightarrow{n \to \infty} \int_{\mathbb{R}^d} f(x) \pi(x) \mathrm{d}x$$

almost surely.

Proof. Corollary 14 ensures that for any $\epsilon > 0$, there are $-\infty < a_1^{(\epsilon)} < a_2^{(\epsilon)} < \infty$ such that $\mathbb{P}(B^{(\epsilon)}) \geq 1 - \epsilon$, where

$$B^{(\epsilon)} := \{\forall n \geq 1, \quad a_1^{(\epsilon)} \leq S_n \leq a_2^{(\epsilon)}\}.$$

Set $K_n^{(\epsilon)} := K^{(\epsilon)} := [a_1^{(\epsilon)}, a_2^{(\epsilon)}]$ for all $n \geq 1$, and construct the truncated process $(\tilde{X}_n^{(\epsilon)}, \tilde{S}_n^{(\epsilon)})$ using these restriction sets in (14). Define $\theta_1^{(\epsilon)} := \phi(a_1^{(\epsilon)}) > 0$ and $\theta_2^{(\epsilon)} := \phi(a_2^{(\epsilon)}) < \infty$.

Let us next verify the above assumptions (A1)–(A4) with some $c \geq 1$, $\beta = 0$ and $V \equiv 1$. The assumption (A1) holds by construction of the process and the Metropolis kernel. For (A2), take $C_n := \mathbb{X}$ for all $n \geq 1$, and notice that $P_s V(x) = 1$ for all $x \in \mathbb{X}$ and $s \in \mathbb{S}$. By Assumption 16 one can estimate for all $s \in K^{(\epsilon)}$ and all $x \in \mathbb{X}$,

$$
\begin{aligned}
P_s(x, A) &\geq \int_A \alpha(x, y) q_s(x - y) \mathrm{d}y \\
&\geq \left(\inf_{x, y \in \mathbb{X}, \, s \in K^{(\epsilon)}} q_s(x - y) \right) \int_A \frac{\pi(y)}{\sup_{z \in \mathbb{X}} \pi(z)} \mathrm{d}y \\
&\geq \theta_2^{-d} \left(\inf_{|z| \leq \mathrm{diam}(\mathbb{X})} \hat{q}(\|\theta_1^{-1} \Sigma^{-1} z\|) \right) c_1 \nu_s(A) \geq \delta \nu_s(A)
\end{aligned}
$$

with a $\delta > 0$, where $\nu_s(A) := \nu(A) := c_1^{-1} \int_A \frac{\pi(y)}{\sup_{z \in \mathbb{X}} \pi(z)} \mathrm{d}y$ and $c_1 > 0$ chosen so that $\nu(\mathbb{X}) = 1$. Assumption 15 ensures that the derivative of ϕ is bounded on the compact set $K_n^{(\epsilon)}$. Therefore, the Frobenius norm $\|\phi(s)\Sigma - \phi(s')\Sigma\| \leq c_2 |s - s'|$ with some finite $c_2(\epsilon)$ and Proposition 27 in Appendix B implies (A3). Finally, it holds that $|H(\tilde{S}_n, \tilde{X}_n, \tilde{Y}_{n+1})| \leq c$, implying (A4).

All (A1)–(A4) hold and $\sum_{k=1}^\infty k^{-1} \eta_k \leq (\sum_{k=1}^\infty k^{-2})^{1/2} (\sum_{k=1}^\infty \eta_k^2)^{1/2} < \infty$, so Theorem 20 yields a strong law of large numbers for the truncated process $\tilde{X}_n^{(\epsilon)}$ in case of a bounded function f. Since $(\tilde{X}_n^{(\epsilon)})_{n \geq 1}$ coincides with the original ASM process $(X_n)_{n \geq 1}$ in $B^{(\epsilon)}$, the strong law of large numbers applies for $X_n(\omega)$ with almost every $\omega \in B^{(\epsilon)}$. Since $\epsilon > 0$ was arbitrary, (16) holds almost surely. $\qquad\square$

Remark 22. Theorem 20 (Theorem 2 of [20]) is a modification of Proposition 6 in [1]. Theorem 21 could be obtained also using other techniques, in particular, the mixingale approach described in [6, 10], or the coupling technique of [15] (resulting in a weak law of large numbers). These other techniques do not, however, apply directly to Theorem 23 below, where Theorem 20 is applied in full strength.

Finally, the second main result considers target densities π with an unbounded support.

Theorem 23. *Suppose π satisfies Assumptions 3 and 11, and the scaling function ϕ satisfies Assumption 15. Assume also that there exist constants $M, \xi < \infty$ such that the function f is bounded by $|f(x)| \leq M \max\{1, e^{\xi \|x\|}\}$ for all $x \in \mathbb{R}^d$. Then, for the ASM process $(X_n, S_n)_{n \geq 1}$ with any $0 < \alpha^* < 1/2$, the strong law of large numbers (16) holds.*

Proof. Proposition 17 ensures that for any $\beta' > 0$ there are a.s. positive Θ_1 and a.s. finite Θ_2 such that

$$
(17) \qquad\qquad\qquad \Theta_1 \leq \phi(S_n) \leq \Theta_2 n^{\beta'}.
$$

Now, similarly as in the proof of Theorem 21, for any $\epsilon > 0$, one can find $0 < \theta_1^{(\epsilon)} \leq \theta_2^{(\epsilon)} < \infty$ such that

$$(18) \qquad \mathbb{P}(\forall n \geq 1 : \theta_1^{(\epsilon)} \leq S_n \leq \theta_2^{(\epsilon)} n^{\beta'}) \geq 1 - \epsilon$$

and construct $(\tilde{X}_n^{(\epsilon)}, \tilde{S}_n^{(\epsilon)})_{n \geq 1}$ using the restriction sets $K_n^{(\epsilon)} := [a_1^{(\epsilon)}, a_2^{(n,\epsilon)}]$, where $\phi(a_1^{(\epsilon)}) = \theta_1^{(\epsilon)}$ and $\phi(a_2^{(\epsilon,n)}) = \theta_2^{(\epsilon)} n^{\beta'}$.

Let $V(x) := c_V \pi^{-1/2}(x)$ with $c_V := \sup_x \pi^{1/2}(x)$. The assumptions (A1) and (A4) hold as verified in the proof of Theorem 21. Proposition 26 in Appendix B with the fact $\det(\theta \Sigma) = \theta^d \det(\Sigma)$ yields (A2) with $\beta = d\beta'$. Assumption 15 ensures that $\phi'(s) \leq c_1 \phi^\kappa(s)$, from which $|\phi(s) - \phi(s')| \leq c_1(\theta_2^{(\epsilon)} n^{\beta'})^\kappa |s - s'| \leq c_2 n^{\kappa \beta'} |s - s'|$ for all $s, s' \in K_n^{(\epsilon)}$. Now, Proposition 27 in Appendix B shows (A3) with $\beta = c_3 \beta'$. To conclude, the assumptions (A1)–(A4) hold with constants (c, β), where $\beta = \beta(\epsilon, \beta') > 0$ can be selected to be arbitrarily small and $c = c(\epsilon, \beta) < \infty$.

In particular, one can let β be sufficiently small to ensure that $\kappa_* \beta < 1/3$ so that $\sum_{k=1}^\infty k^{\kappa_* \beta - 1} \eta_k < \infty$ as in the proof of Theorem 21. One can take $\gamma = 1/2$ and observe that $V^\gamma(x) \geq c_4 \max\{1, e^{c_5 \|x\|^{\rho - 1}}\}$ for some $c_4, c_5 > 0$ implying that $\sup_x |f(x)|/V^\gamma(x) < \infty$. Theorem 20 ensures that the strong law of large numbers holds in the set (18), and a.s. by letting $\epsilon \to 0$. $\square$

Remark 24. It is possible to extend Theorems 21 and 23 to an algorithm using the Adaptive Metropolis algorithm within the ASM framework [3, 5] by applying the observations in Remark 19.

Acknowledgements

References

[1] C. Andrieu and É. Moulines. On the ergodicity properties of some adaptive MCMC algorithms. *Ann. Appl. Probab.*, 16(3):1462–1505, 2006.

[2] C. Andrieu and C. P. Robert. Controlled MCMC for optimal sampling. Technical Report Ceremade 0125, Université Paris Dauphine, 2001.

[3] C. Andrieu and J. Thoms. A tutorial on adaptive MCMC. *Statist. Comput.*, 18 (4):343–373, Dec. 2008.

[4] C. Andrieu, É. Moulines, and P. Priouret. Stability of stochastic approximation under verifiable conditions. *SIAM J. Control Optim.*, 44(1):283–312, 2005.

[5] Y. Atchadé and G. Fort. Limit theorems for some adaptive MCMC algorithms with subgeometric kernels. *Bernoulli*, 16(1):116–154, Feb. 2010.

[6] Y. F. Atchadé and J. S. Rosenthal. On adaptive Markov chain Monte Carlo algorithms. *Bernoulli*, 11(5):815–828, 2005.

[7] M. Bédard. Optimal acceptance rates for Metropolis algorithms: Moving beyond 0.234. *Stochastic Process. Appl.*, 118(12):2198–2222, 2008.

[8] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice.* Chapman & Hall/CRC, Boca Raton, Florida, 1998. ISBN 0-412-05551-1.

[9] W. R. Gilks, G. O. Roberts, and S. K. Sahu. Adaptive Markov chain Monte Carlo through regeneration. *J. Amer. Statist. Assoc.*, 93(443):1045–1054, 1998.

[10] H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001.

[11] P. Hall and C. C. Heyde. *Martingale Limit Theory and Its Application.* Academic Press, New York, 1980. ISBN 0-12-319350-8.

[12] S. F. Jarner and E. Hansen. Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.*, 85:341–361, 2000.

[13] E. Nummelin. MC's for MCMC'ists. *International Statistical Review*, 70(2):215–240, 2002.

[14] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods.* Springer-Verlag, New York, 1999. ISBN 0-387-98707-X.

[15] G. O. Roberts and J. S. Rosenthal. Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *J. Appl. Probab.*, 44(2):458–475, 2007.

[16] G. O. Roberts and J. S. Rosenthal. Examples of adaptive MCMC. *J. Comput. Graph. Statist.*, 18(2):349–367, 2009.

[17] G. O. Roberts and J. S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:20–71, 2004.

[18] G. O. Roberts and J. S. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statist. Sci.*, 16(4):351–367, 2001.

[19] G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, 7(1):110–120, 1997.

[20] E. Saksman and M. Vihola. On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. *Ann. Appl. Probab.*, to appear. Preprint, arXiv:0806.2933v3.

[21] C. Sherlock and G. Roberts. Optimal scaling of the random walk Metropolis on elliptically symmetric unimodal targets. *Bernoulli*, 15(3):774–798, 2009.

[22] S.-D. Wang, T.-S. Kuo, and C.-F. Hsu. Trace bounds on the solution of the algebraic matrix Riccati and Lyapunov equation. *IEEE Trans. Automat. Control*, 31(7):654–656, July 1986.

Appendix A. Half-Space Approximation

Proof of Lemma 13. Fix an $\epsilon' > 0$. By the uniform smoothness of $\{\partial A_i\}_{i \in I}$, one can let $\delta > 0$ be so small that $\|n_i(y) - n_i(z)\| \leq \epsilon'$ for all $i \in I$ and $y, z \in \partial A_i$ with $\|y - z\| \leq 2\delta$.

Fix an $i \in I$, an $x \in A_i$ and a $r \in [0, \delta]$. If $\overline{B}(x, r) \setminus A_i = \emptyset$, one can let T be any half-space passing through x. Suppose for the rest of the proof that $\overline{B}(x, r) \setminus A_i \neq \emptyset$
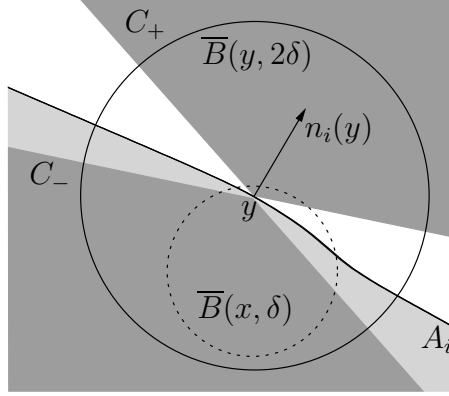
FIGURE 1. Illustration of the half-space approximation. The set A_i is shown in light grey, and the cones C_- and C_+ in dark grey.

and let $y \in \overline{B}(x, r) \cap \partial A_i$. Consider the open cones

$$C_- := \{y + z : n_i(y) \cdot z < -\epsilon' \|z\|\}$$
$$C_+ := \{y + z : n_i(y) \cdot z > \epsilon' \|z\|\}$$

illustrated in Figure 1. We shall verify that $\overline{B}(y, 2\delta) \cap C_- \subset \overline{B}(y, 2\delta) \cap A_i$ and $\overline{B}(y, 2\delta) \cap C_+ \subset \overline{B}(y, 2\delta) \setminus A_i$.

Namely, let $u \in \overline{B}(y, 2\delta) \cap C_-$ and write $u = y + z$. Suppose that $u \notin A_i$ and define $t_0 := \inf\{t \in [0, 1] : y + tz \notin A_i\}$. Let $u_0 := y + t_0 z$ and notice that $u_0 \in \overline{B}(y, 2\delta) \cap \partial A_i$. Moreover, the line segment $y + tz$ with $t \in [0, 1]$ passes through ∂A_i at u_0 and therefore $n_i(u_0) \cdot z \geq 0$, since n_i is the outer-pointing normal of A_i. On the other hand,

$$n_i(u_0) \cdot \frac{z}{\|z\|} = (n_i(u_0) - n_i(y)) \cdot \frac{z}{\|z\|} + n_i(y) \cdot \frac{z}{\|z\|}$$
$$< \|n_i(u_0) - n_i(y)\| - \epsilon' < 0,$$

which is a contradiction, implying $C_- \cap \overline{B}(y, 2\delta) \subset A_i \cap \overline{B}(y, 2\delta)$. The case with C_+ is verified similarly.

Let us define the half-space $T := \{y - 2\epsilon' r n_i(y) + z : z \cdot n_i(y) < 0\}$. It holds that $\overline{B}(y, 2r) \cap T \subset \overline{B}(y, 2r) \cap C_-$ since taking $y + w \in \overline{B}(y, 2r) \cap T$ one has $n_i(y) \cdot w < -2\epsilon' r \leq -\epsilon' \|w\|$. On the other hand, $\overline{B}(y, 2r) \cap C_- \subset \overline{B}(y, 2r) \cap A_i$ and $\overline{B}(x, r) \subset \overline{B}(y, 2r)$, so $\overline{B}(x, r) \cap T \subset \overline{B}(x, r) \cap A_i$. Clearly, $d(y, T) = 2\epsilon' r$, and since $x \notin C_+$ one has $n_i(y) \cdot (x - y) \leq \epsilon' \|x - y\| \leq \epsilon' r$. To conclude, $d(x, T) \leq 3\epsilon' r$, and taking $\epsilon' = \epsilon/3$ yields the claim. $\qquad\square$

APPENDIX B. SIMULTANEOUS PROPERTIES FOR METROPOLIS KERNELS

Let us define the following generalisation of Assumption 16.

Assumption 25. Let $\mathcal{C}_d \subset \mathbb{R}^{d \times d}$ stand for the symmetric and positive definite matrices. Suppose $\mathcal{P} \subset \mathcal{C}_d$ and $\{q_s\}_{s \in \mathcal{P}}$ is a family of probability densities defined through

$$(19) \qquad\qquad\qquad q_s(z) := |\det(s)|^{-1} \hat{q}(\|s^{-1}z\|),$$

where $\hat{q} : [0, \infty) \to (0, \infty)$ is a bounded, decreasing, and differentiable function, satisfying the conditions in Assumption 16. Moreover, suppose that there is a $\kappa > 0$ such that the eigenvalues of each $s \in \mathcal{P}$ are bounded from below by κ.

Proposition 26. *Suppose π satisfies Assumption 3 and the family $\{q_s\}_{s \in \mathcal{P}}$ satisfies Assumption 25 with some $\kappa > 0$. Let P_s be the Metropolis transition probability defined in* (6) *and using the proposal density q_s. Then, there exists a compact set $C \subset \mathbb{R}^d$, a probability measure ν on C and a constant $b \in [0, \infty)$ such that for all $s \in \mathcal{P}$, $x \in \mathbb{R}^d$ and measurable $A \subset \mathbb{R}^d$,*

$$(20) \qquad\qquad\qquad P_s V(x) \quad\leq\quad \lambda_s V(x) + b \mathbb{1}_C(x)$$

$$(21) \qquad\qquad\qquad P_s(x, A) \quad\geq\quad \delta_s \mathbb{1}_C(x) \nu(A)$$

where $V(x) := c_V \pi^{-1/2}(x) \geq 1$ with $c_V := \sup_x \pi^{1/2}(x)$ and the constants $\lambda_s, \delta_s \in (0, 1)$ satisfy the bound

$$(1 - \lambda_s)^{-1} \vee \delta_s^{-1} \leq c |\det(s)|^{-1}$$

for some constant $c \geq 1$.

Proof. Proposition 26 is a generalisation of [20, Proposition 18] considering Gaussian densities q_s. We shall describe the changes that are needed in the proof of [20, Proposition 18].

Let $s \in \mathcal{P}$. For a non-negative function f, one can write by Fubini's theorem

$$\int_{\mathbb{R}^d} f(z + x) q_s(z) \mathrm{d}z = |\det(s)|^{-1} \int_0^{\hat{q}(0)} \int_{\{\hat{q}(\|s^{-1}z\|) \geq t\}} f(z + x) \mathrm{d}z \mathrm{d}t$$

$$= -|\det(s)|^{-1} \int_0^\infty \int_{E_u} f(y) \mathrm{d}y \hat{q}'(u) \mathrm{d}u$$

where the substitution $t = \hat{q}(u)$ was used, and $E_u := \{x + z : \|s^{-1}z\| \leq u\}$. One has $\|s^{-1}z\| \leq \kappa^{-1}\|z\|$, and thus $E_u \supset \overline{B}(x, u\kappa)$. The conditions in Assumption 16 for the derivative $\hat{q}'$ corresponds to the estimate obtained in [20, Lemma 17] for a Gaussian family, that is, $\hat{q} = e^{-x^2/2}$.

These facts are enough to complete the proof of [20, Proposition 18] to yield the claim. $\qquad\square$

Proposition 27. *Suppose the family $\{q_s\}_{s \in \mathcal{P}}$ satisfies Assumption 25 with some $\kappa > 0$. Suppose, in addition, that either*

(i) $V \equiv 1$ or

(ii) π satisfies Assumption 3 and $V(x) := c_V \pi^{-1/2}(x) \geq 1$ with $c_V := \sup_x \pi^{1/2}(x)$.

Then, there are constants $c_1, c_2 > 0$ such that for the Metropolis transition probability P_s given in (6)*, it holds that*

$$(22) \qquad\qquad \|P_s f - P_{s'} f\|_{V^r} \leq c_1 \max\{\|s\|, \|s'\|\}^{c_2} \|f\|_{V^r} \|s - s'\|$$

for all $s, s' \in \mathcal{P}$ and $r \in [0, 1]$. The matrix norm above is the Frobenius norm defined as $\|a\| := \sqrt{\mathrm{tr}(a^T a)}$.

Proof. Consider first (i). From the definition of the Metropolis kernel (6), one obtains

$$\sup_x |P_s f(x) - P_{s'} f(x)| \le 2 \sup_x |f(x)| \int_{\mathbb{X}} |q_s(x) - q_{s'}(x)| \mathrm{d}x.$$

For (ii), Proposition 12 of [1] shows that for any $r \in [0, 1]$ it holds that

$$\|P_s f - P_{s'} f\|_{V^r} \le 2 \|f\|_{V^r} \int_{\mathbb{X}} |q_s(x) - q_{s'}(x)| \mathrm{d}x$$

so it is sufficient to consider only the total variation of the proposal distributions.

As in [10] and [1], one can write

$$\int_{\mathbb{X}} |q_s(x) - q_{s'}(x)| \mathrm{d}x = \int_{\mathbb{X}} \left| \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} q_{s_t}(x) \mathrm{d}t \right| \mathrm{d}x$$

where $s_t := s' + t(s - s')$. Let us compute

$$\frac{\mathrm{d}}{\mathrm{d}t} q_{s_t}(x) = |\det(s_t)|^{-1} \left(- \mathrm{tr}\left(s_t^{-1}(s - s') \right) q_{s_t}(x) + \hat{q}'(\|s_t^{-1}x\|) \frac{\mathrm{d}}{\mathrm{d}t} \|s_t^{-1}x\| \right)$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t} \|s_t^{-1}x\| = - \left(\frac{s_t^{-1}x}{\|s_t^{-1}x\|} \right)^T s_t^{-1}(s - s') s_t^{-1}x.$$

Since $s - s'$ and s_t^{-1} are symmetric and s_t^{-1} positive definite, it holds that $|\mathrm{tr}\left(s_t^{-1}(s - s') \right)| \le \mathrm{tr}(s_t^{-1}) \max_{1 \le i \le d} |\lambda_i| \le \mathrm{tr}(s_t^{-1})\|s - s'\|$ where λ_i are the eigenvalues of $s - s'$ [see, e.g, 22]. Since the Frobenius norm is sub-multiplicative,

$$\int_{\mathbb{X}} |q_s(x) - q_{s'}(x)| \mathrm{d}x$$

$$\le \sup_{t \in [0,1]} |\det(s_t)|^{-1} \left(\mathrm{tr}(s_t^{-1}) + \|s_t^{-1}\|^2 \int_{\mathbb{X}} \|x\| \left| \hat{q}'(\|s_t^{-1}x\|) \right| \mathrm{d}x \right) \|s - s'\|$$

$$\le \kappa^{-d} \left(d\kappa^{-1} + d\kappa^{-2} c_d \sup_{\|u\|=1, t \in [0,1]} \int_0^\infty r^d |\hat{q}'(r\|s_t^{-1}u\|)| \mathrm{d}r \right) \|s - s'\|$$

by polar integration. Denote $\lambda = \lambda(u, t) := \|s_t^{-1}u\|$, and observe that since $\hat{q}$ is decreasing, integration by parts yields

$$\int_0^M r^d |\hat{q}'(\lambda r)| \mathrm{d}r = \frac{d}{\lambda} \int_0^M r^{d-1} \hat{q}(\lambda r) \mathrm{d}r - M^d \frac{q(\lambda M)}{\lambda}$$

$$\le \frac{d}{\lambda^{d+1}} \int_0^\infty u^{d-1} \hat{q}(u) \mathrm{d}u = \frac{dc_{\hat{q}}}{\lambda^{d+1}}$$

for all $M > 0$. Since λ^{-1} is smaller, for any $\|u\| = 1$ and $t \in [0,1]$, than the maximum eigenvalue of s and s', which is smaller than $\max\{\|s\|, \|s'\|\}$, we obtain

$$\int_{\mathbb{X}} |q_s(x) - q_{s'}(x)| \mathrm{d}x \le c_1 \max\{\|s\|, \|s'\|\}^{d+1} \|s - s'\|$$

concluding the proof. $\qquad\qquad\square$

Lemma 28. *Suppose the template proposal density q is given as $q(z) = c\hat{q}(\|\Sigma^{-1}z\|)$ where $c > 0$ is a constant and $\Sigma \subset \mathbb{R}^{d \times d}$ is a symmetric and positive definite matrix, and*

(i) $\hat{q}(x) = e^{-x^2/2}$, or
(ii) $\hat{q}(x) = (1 + x^2)^{-d/2-\gamma}$ for some $\gamma > 0$.

That is, q is a (multivariate) Gaussian or Student distribution, respectively. Then, q satisfies Assumption 16.

Proof. For (i), Assumption 16 is implied by [20, Lemma 17].

Assume then that $\hat{q}$ has the form (ii) and fix an $\epsilon > 0$. By the mean value theorem, denoting $c_1 := d + 2\gamma$ and $\alpha := d/2 + \gamma + 1$, one can write for some $\epsilon' \in [0, \epsilon]$

$$
\begin{aligned}
\hat{q}'(x) - 2\hat{q}'(x + \epsilon) &= c_1 x \left(\frac{2}{(1 + (x+\epsilon)^2)^\alpha} - \frac{1}{(1 + x^2)^\alpha} \right) \\
&= c_1 x \left(\frac{1}{(1 + (x+\epsilon)^2)^\alpha} - \frac{2\alpha\epsilon(x+\epsilon')}{(1 + (x+\epsilon')^2)^{\alpha+1}} \right) \\
&\ge \frac{c_1 x}{(1 + (x+\epsilon)^2)^\alpha} \left(1 - 2\alpha\epsilon \left(\frac{1 + (x+\epsilon)^2}{1 + (x+\epsilon')^2} \right)^\alpha \right) > 0
\end{aligned}
$$

for all $x > 0$, whenever $\epsilon > 0$ is sufficiently small. $\qquad\qquad\square$

Matti Vihola, Department of Mathematics and Statistics, University of Jyväskylä, P.O.Box 35 (MaD), FI-40014 University of Jyväskylä, Finland
  E-mail address: matti.vihola@iki.fi
  URL: http://iki.fi/mvihola/